

QFQ - Feature #10715

wkhtmltopdf ersetzen: puppeteer

04.06.2020 19:41 - Marc Egger

Status:	New	Start date:	08.06.2021
Priority:	High	Due date:	
Assignee:	Benjamin Baer	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	21.7.0	Spent time:	0.00 hour
Discuss:			
Description			
Prio4 von Nicola			
Varianten:			
<ul style="list-style-type: none">• puppeteer https://pptr.dev/<ul style="list-style-type: none">◦ wkhtmltopdf VS puppeteer: https://medium.com/collectiv-stories/puppeteer-vs-wkhtmltopdf-and-why-i-created-a-new-module-9466eb1db7d1◦ https://crpdf.evens.eu/ (alt)• Pandoc• generate latex using a twig template (actually a separate feature)• Chrome Headless: https://developers.google.com/web/updates/2017/04/headless-chrome#create_a_pdf_dom			
Tools			
<ul style="list-style-type: none">• Artikel zum Thema: https://www.paperplane.app/blog/modern-html-to-pdf-conversion-2019/			
Diese koennen vermutlich alle kein JS, sind dafuer aber als LIB oder API sehr schnell:			
<ul style="list-style-type: none">• https://weasyprint.org/• https://www.pagedjs.org• https://github.com/esbenp/pdf-bot			

Associated revisions

Revision 17f774fd - 29.06.2021 22:41 - Carsten Rose

WIP. Refs #10715

History

#1 - 04.06.2020 19:44 - Marc Egger

- Description updated

#2 - 10.06.2020 09:07 - Carsten Rose

- Subject changed from wkhtmltopdf ersetzen to wkhtmltopdf ersetzen: puppeteer

#4 - 07.06.2021 11:04 - Carsten Rose

- Priority changed from Normal to High

- Target version changed from 21.8.0 to 21.7.0

#5 - 07.06.2021 13:24 - Carsten Rose

- Description updated

#6 - 07.06.2021 13:29 - Carsten Rose

- Description updated

#7 - 07.06.2021 15:24 - Carsten Rose

Test

- Test von Fabric: <http://webwork16.math.uzh.ch/crose/qfq/index.php?id=annotatfabric&form=annotateGraphic&r=1>
- Test von Calendar: <http://webwork16.math.uzh.ch/crose/qfq/index.php?id=96>

Umgebung

- webwork16 (Ubuntu 18):
- crpdf als linux Binary installiert: <https://github.com/JorgenEvans/crPDF/>

```
# Test als npm - Installation hat funktioniert, aber das Binary blieb haengen
# FALSCH: CR hat einfach nur nicht lange genug gewartet. Beim ersten Start wird in das Home des users ein komp
letter Chrome geladen!!! >300MB
```

```
[root@host]
$ npm install -g crpdf
```

```
# Aufruf bleibt haengen
$ crpdf https://math.uzh.ch/final.pdf
```

```
# Installation des fertigen Linux Pakets https://github.com/JorgenEvans/crPDF/releases
[root@host] mv /var/tmp/crpdf /usr/local/bin/crpdf-pkg
```

```
# Aufruf OK fuer root UND fuer www-data
[root@host] cd; crpdf-pkg 'http://webwork16.math.uzh.ch/crose/qfq/index.php?id=annotatfabric&form=annotateGra
phic&r=1' final.pdf
[www-data@host] cd; crpdf-pkg 'http://webwork16.math.uzh.ch/crose/qfq/index.php?id=annotatfabric&form=annotat
eGraphic&r=1' final.pdf
```

#8 - 07.06.2021 15:30 - Carsten Rose

- File final.pdf added

#9 - 07.06.2021 15:50 - Carsten Rose

- Assignee changed from Carsten Rose to Benjamin Baer
- Start date changed from 04.06.2020 to 08.06.2021

Liste der wkhtml Optionen: <https://wkhtmltopdf.org/usage/wkhtmltopdf.txt>

Ziel: zumindest die aktuell verwendeten Optionen sollten funktionieren.

Wichtig sind die Cookies. **Achtung** : in dem --cookie-jar File werden alle aktuellen Cookies kopiert (QFQ, FE, PHP, ...)

```
user arguments
-----
```

```
--margin-bottom=20mm
--footer-left="Seite: [page]/[toPage]"
--footer-right="1234, UZH Doc.Mobility"
--footer-font-size=8
--footer-spacing=10
```

```
QFQ added arguments
-----
```

```
--cookie-jar '/tmp/qfq.cookie...'
'--custom-header User-Agent ' . escapeshellarg($_SERVER['HTTP_USER_AGENT'] ?? '')
--custom-header-propagation
```

#11 - 08.06.2021 19:01 - Benjamin Baer

Damit ich mal wieder meine Browser Fenster aufräumen kann, hier noch einige Punkte zum jetzigen Status.

Unser crpdf: <https://git.math.uzh.ch/utilities/crpdf-custom>

Download builds: <https://www.math.uzh.ch/repo/?dir=crpdf>

About Cookies:

Bug in old puppeteer: <https://github.com/puppeteer/puppeteer/issues/717>

Cookies and Puppeteer uebergeben: <https://stackoverflow.com/questions/50418994/pass-signed-cookie-to-puppeteer>

Multiple Cookies: <https://stackoverflow.com/questions/50584770/passing-multiple-cookies-to-puppeteer>

Spread Syntax: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Operators/Spread_syntax

Set Cookie has no effect: <https://github.com/puppeteer/puppeteer/issues/1342>

Browserfetcher problems: Newer versions of browserFetcher don't support the current getStandardVersion anymore. Puppeteer v10 uses Chromium 92.0.4512.0, see top of: <https://github.com/puppeteer/puppeteer/blob/v10.0.0/docs/api.md>
To get the slug to download the version (eg 885287 for linux chrome 92.0.4512.0), go to: <http://omahaproxy.appspot.com/>
Download brings error that it wasn't found with 885287
<https://github.com/puppeteer/puppeteer/blob/v10.0.0/docs/api.md#class-browserfetcher>

About NPM:

cmd: <https://nodejs.org/en/knowledge/command-line/how-to-parse-command-line-arguments/>

#12 - 10.06.2021 07:52 - Carsten Rose

- File `qfq.cookie.8ZORgh` added

#13 - 15.06.2021 17:26 - Carsten Rose

- crpdf (Puppeteer) von Benj: <https://git.math.uzh.ch/utilities/crPDF/-/tree/master>
- Das json Array muss wie folgt aussehen:

```
{ "cookies": [
  {
    "name": "test",
    "value": "3fh822ji4f244",
    "domain": "w16.math.uzh.ch",
    "path": "/my/"
  },
  {
    "name": "mondieux",
    "value": "oh mon dieux",
    "domain": "w16.math.uzh.ch",
    "path": "/my/"
  }
]}
```

Files

final.pdf	59.4 KB	07.06.2021	Carsten Rose
qfq.cookie.8ZORgh	1 KB	10.06.2021	Carsten Rose